



ORIGINAL RESEARCH PAPER

Predicting Corn Yield Using Ensemble Learning with Satellite and Meteorological Data

A. Sabzali Yameqani*, A. A. Alesheikh

Department of Geospatial Information Systems, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran

ABSTRACT

Received: 31 August 2024
 Reviewed: 23 September 2024
 Revised: 22 October 2024
 Accepted: 03 December 2024

KEYWORDS:

Ensemble Learning
 Yield Prediction
 Pelican Optimization Algorithm
 Corn

* Corresponding author

✉ asabzali@mail.kntu.ac.ir

☎ (+9821) 88786212

Background and Objectives: Precise agricultural yield prediction is among the most important tools for managing agricultural resources, improving food security, and enhancing the productivity of international trade in agricultural products. The satellite remote sensing images has become widely adopted because traditional methods cannot provide the needed accurate and timely predictions, and it covers large areas while providing accurate data. The advances in machine learning and ensemble learning have identified the complex interaction of environmental variables with crop yield. In modern times, ensemble learning models have achieved much higher prediction accuracy and provided useful insights to farmers and policy makers.

This study aims to develop an innovative model that combines the XGBoost algorithm with the Pelican Optimization Algorithm (POA) to predict corn yields more accurately in the U.S. Midwest. The approach will provide an opportunity for the pre-harvest yield prediction by considering the plant phenological stages and optimal time range from July to August. The model will help the decision-makers to take effective measures on resource management to overcome the climate fluctuations and develop better agricultural policies.

Methods and Materials: This research focuses on predicting corn yields in five key corn-producing states in the U.S. Midwest (Illinois, Iowa, Minnesota, North Dakota, and South Dakota). This paper will utilize remote sensing information, including NDVI (Normalized Difference Vegetation Index), EVI (Enhanced Vegetation Index), LAI (Leaf Area Index), FPAR (Fraction of Photosynthetically Active Radiation), GPP (Gross Primary Production), and ET (Evapotranspiration); meteorological data, including temperature and precipitation; cropland data; and yield statistics during the growing season over the period 2011–2020 (May to September). XGBoost ensemble learning was used, whose hyperparameters were optimized with the Pelican Optimization Algorithm (POA) to enhance its accuracy. Filtration was performed on data using the VFI index. Nine years were used as training data, while one year was used as a test. For evaluating the performance, MAPE, MBE, MAE, RMSE, and the correlation coefficient have been used.

Findings: The evaluation results of the POA-XGBoost model demonstrated its outstanding performance in predicting corn yields. During the 2011–2020 timeframe, validation trends highlighted variations in prediction accuracy and bias. In the first period, which includes 2011–2014, the errors went down and the prediction accuracy improved: MAPE reached 6.26%, while in 2014 the correlation coefficient increased to 0.9372. During the middle period of 2015–2018, the errors and positive bias showed an upward trend, especially during 2018, where MBE rose to 0.8039 and the correlation coefficient fell to 0.8083. However, the last two years (2019–2020) revealed much improved results: MAPE comprises 6.57%, while the correlation coefficient is as high as 0.9237 in 2020.

Conclusion: The optimized POA-XGBoost model demonstrated high capability in predicting corn yields under diverse climatic conditions and can be extended to forecast other crops in the future. Advanced ensemble learning techniques combined with diverse data sources, such as satellite imagery and meteorological data, provide effective solutions for improving crop yield predictions. The study calls for the development of new hybrid models that will enable farmers and managers to better manage resources, increase productivity, and minimize risks.



NUMBER OF REFERENCES

32



NUMBER OF FIGURES

2



NUMBER OF TABLES

5

مقاله پژوهشی

پیش‌بینی میزان محصول ذرت با بهره‌گیری از یادگیری گروهی، داده‌های ماهواره و هواشناسی

علی سبزعلی یمقانی^{*}، علی اصغر آل شیخ

گروه سیستم‌های اطلاعات مکانی، دانشکده مهندسی نقشه‌برداری، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

چکیده

پیشینه و اهداف: پیش‌بینی دقیق میزان محصول کشاورزی یکی از ابزارهای اساسی برای مدیریت منابع کشاورزی، بهبود امنیت غذایی و افزایش بهره‌وری تجارت بین‌المللی محصولات است. با توجه به محدودیت‌های روش‌های سنتی در پیش‌بینی دقیق و سریع، تصاویر سنجنش از دور ماهواره‌ای به دلیل توانایی پوشش مناطق وسیع و ارائه داده‌های دقیق، به‌طور گسترده‌ای جایگزین شده‌اند. در این میان، پیشرفت‌های یادگیری ماشین و یادگیری گروهی نقش مهمی در شناسایی روابط پیچیده میان متغیرهای محیطی و بازده محصولات ایفا کرده‌اند. مدل‌های نوین یادگیری گروهی، توانسته‌اند دقت پیش‌بینی‌ها را به‌طور قابل توجهی افزایش دهند و اطلاعات ارزشمندی برای کشاورزان و سیاست‌گذاران فراهم کنند. این پژوهش با هدف توسعه مدلی نوآورانه مبتنی بر ترکیب مدل XGBoost و الگوریتم بهینه‌سازی پلیکان (POA) به‌منظور پیش‌بینی دقیق‌تر میزان محصول ذرت در غرب میانه ایالات متحده انجام شده است. رویکرد پیشنهادی با بررسی تأثیر بازه‌های مختلف فنولوژی گیاه و انتخاب بازه بهینه (ژوئیه تا اوت)، امکان پیش‌بینی دقیق محصول را پیش از برداشت فراهم می‌کند. این مدل می‌تواند تصمیم‌گیری‌های مؤثری در زمینه مدیریت منابع، کاهش خسارات ناشی از نوسانات اقلیمی، و ارتقای سیاست‌های کشاورزی ارائه دهد.

تاریخ دریافت: ۱۰ شهریور ۱۴۰۳
تاریخ داوری: ۰۲ مهر ۱۴۰۳
تاریخ اصلاح: ۰۱ آبان ۱۴۰۳
تاریخ پذیرش: ۱۳ آذر ۱۴۰۳

واژگان کلیدی:

یادگیری گروهی
پیش‌بینی میزان محصول
الگوریتم بهینه‌سازی پلیکان
ذرت

روش و مواد: این پژوهش با هدف پیش‌بینی میزان محصول ذرت در پنج ایالت غرب میانه کشور آمریکا (ایلینوی، آیووا، مینه‌سوتا، داکوتای شمالی و داکوتای جنوبی) که از مناطق اصلی تولید ذرت هستند، انجام شده است. داده‌های استفاده‌شده شامل اطلاعات سنجنش از دور (مانند شاخص نرمال شده تفاوت پوشش گیاهی (NDVI)، شاخص پوشش گیاهی بهبود یافته (EVI)، شاخص سطح برگ (LAI)، شاخص جزء جذب‌شده تابش فتوسنتز فعال (FPAR)، تولید اولیه ناخالص (GPP)، و تبخیر و تعرق (ET))، داده‌های آب‌وهوایی (نظیر دما و بارش)، داده‌های زمین‌های زراعی، و بازده محصول، طی بازه زمانی ۲۰۱۱ تا ۲۰۲۰ و فصل رشد (اواسط اردیبهشت تا اواسط مهر) جمع‌آوری شده‌اند. مدل‌سازی با استفاده از الگوریتم یادگیری گروهی XGBoost انجام شد و تنظیم پارامترهای آن با استفاده از الگوریتم بهینه‌سازی پلیکان (POA) برای بهبود دقت صورت گرفت. داده‌ها ابتدا با شاخص VFI فیلتر شده و از ۹ سال به‌عنوان داده آموزشی و یک سال به‌عنوان داده آزمون استفاده شد. عملکرد مدل با شاخص‌هایی نظیر MAPE، MAE، MBE، RMSE و ضریب همبستگی ارزیابی شد.

یافته‌ها: نتایج ارزیابی مدل POA-XGBoost نشان داد که این روش عملکردی برجسته در پیش‌بینی میزان محصول ذرت دارد. در بازه زمانی ۲۰۱۱ تا ۲۰۲۰، روند اعتبارسنجی پیش‌بینی میزان محصول ذرت نشان‌دهنده تغییرات در دقت پیش‌بینی‌ها و بایاس بود. در سال‌های ابتدایی (۲۰۱۱ تا ۲۰۱۴)، مقادیر خطاها کاهش و دقت پیش‌بینی‌ها بهبود یافت؛ به‌طوری‌که در سال ۲۰۱۴ مقدار MAPE به ۶/۲۶٪ و ضریب همبستگی به ۰/۹۳۷۲ رسید. در سال‌های میانی (۲۰۱۵ تا ۲۰۱۸)، افزایش خطاها و بایاس مثبت به همراه کاهش دقت پیش‌بینی‌ها مشاهده شد، به‌ویژه در سال ۲۰۱۸ که مقدار MBE به ۰/۸۰۳۹ و ضریب همبستگی به ۰/۸۰۸۳ کاهش یافت. در سال‌های پایانی (۲۰۱۹ و ۲۰۲۰)، بهبود قابل توجهی در دقت پیش‌بینی‌ها مشاهده شد؛ به‌طوری‌که در سال ۲۰۲۰ مقدار MAPE به ۶/۵۷٪ و ضریب همبستگی به ۰/۹۲۳۷ رسید.

نتیجه‌گیری: مدل بهینه POA-XGBoost توانایی بالایی در پیش‌بینی بازده محصول ذرت در شرایط مختلف آب‌وهوایی نشان داده است و می‌تواند در آینده برای پیش‌بینی سایر محصولات کشاورزی نیز گسترش یابد. استفاده از تکنیک‌های پیشرفته یادگیری گروهی و داده‌های متنوع مانند تصاویر ماهواره‌ای و داده‌های هواشناسی، راه‌حل‌های مؤثری برای بهبود پیش‌بینی بازده محصولات ارائه می‌دهد. این پژوهش بر ضرورت توسعه مدل‌های ترکیبی جدید تأکید دارد که می‌توانند به کشاورزان و مدیران در مدیریت منابع، افزایش بهره‌وری، و کاهش ریسک کمک کنند.

^{*} نویسنده مسئول
asabzali@mail.kntu.ac.ir
۰۲۱-۸۸۷۸۶۲۱۲

مقدمه

ذرت به عنوان یکی از مهم‌ترین مواد غذایی اصلی جهان، نقش حیاتی در تأمین کالری روزانه بیش از ۴.۵ میلیارد نفر ایفا می‌کند. این محصول به دلیل محتوای کالری بالا و سازگاری با شرایط زیست‌محیطی از اهمیت ویژه‌ای برخوردار است و انتظار می‌رود با افزایش جمعیت جهانی، تقاضای آن در سال‌های آتی به طور چشمگیری افزایش یابد [۱، ۲]. برآورد دقیق میزان محصول، به‌ویژه برای مسائلی نظیر مدیریت کشاورزی، سیاست‌گذاری ملی مواد غذایی و تجارت بین‌المللی محصولات زراعی، امری حیاتی محسوب می‌شود. روش‌های پیش‌بینی میزان محصول، به‌ویژه کاربرد تصاویر ماهواره‌ای و تکنیک‌های سنجش از دور، به دلیل توانایی در پوشش مناطق وسیع و ارائه ارزیابی دقیق‌تر از بازده محصولات، اهمیت روزافزونی یافته‌اند [۳-۵]. برآورد عملکرد محصولات کشاورزی همچنین به کشاورزان، سازمان‌های بهبود تجاری و نهادهای دولتی امکان می‌دهد تا تصمیمات آگاهانه‌ای در جهت ارتقای سیاست‌های غذایی و تجارت بین‌المللی اتخاذ کنند [۱، ۶-۱۰].

مدل‌های پیش‌بینی بازده محصولات زراعی نقش کلیدی در مدیریت کشاورزی و امنیت غذایی ایفا می‌کنند. مدل‌های مبتنی بر فرآیند فیزیکی و یادگیری ماشینی دو روش متداول در این زمینه هستند. مدل‌های مبتنی بر فرآیند به دلیل اعتبار علمی و قابلیت تفسیر بالا بسیار ارزشمندند، اما نیاز به داده‌های دقیق حسگرهای میدانی و محلی، کاربرد آن‌ها را در مقیاس بزرگ محدود می‌کند. در مقابل، مدل‌های یادگیری ماشینی با استفاده از داده‌های تاریخی بازده محصول و شاخص‌های مرتبط، بدون نیاز به نصب حسگرهای مختلف، پیش‌بینی‌های کارآمدتری ارائه می‌دهند [۱۱]. این مدل‌ها با توانایی استخراج روابط پیچیده، به دلیل پیشرفت‌های سریع در فناوری، به‌طور گسترده‌ای در مقیاس‌های بزرگ استفاده شده‌اند. نمونه‌هایی از این روش‌ها شامل جنگل تصادفی، رگرسیون بردار پشتیبان، شبکه‌های عصبی پیچشی و بازگشتی، حافظه طولانی کوتاه‌مدت و XGBoost است که در پیش‌بینی بازده محصول موفقیت‌های چشمگیری داشته‌اند [۱۲، ۱۳].

تحقیقات متعددی در زمینه پیش‌بینی میزان محصولات کشاورزی با استفاده از تکنیک‌های سنجش از دور و یادگیری ماشینی انجام شده است. این روش‌ها به کشاورزان و محققان کمک می‌کنند تا با دقت بیشتری میزان محصولات را پیش‌بینی کنند و بدین ترتیب تصمیم‌گیری‌های بهتری در زمینه مدیریت کشاورزی اتخاذ نمایند. Kim و Lee [۱۴] با استفاده از چهار روش یادگیری ماشینی مختلف (ماشین بردار پشتیبان، جنگل تصادفی، ERT (Extremely Randomized Trees) و یادگیری عمیق) در ایالت آیووا دریافتند که روش یادگیری عمیق بیشترین دقت را در پیش‌بینی میزان محصول ذرت دارد. آن‌ها سه گروه دوره‌ای مختلف برای بررسی حساسیت‌های فصلی بازده ذرت تنظیم کردند و دریافتند که ترکیب بهینه ماه‌ها می‌تواند در مدل‌سازی آماری میزان محصول موثر باشد.

مطالعات مختلف نشان داده‌اند که تکنیک‌های یادگیری ماشینی و یادگیری عمیق به طور قابل‌توجهی دقت پیش‌بینی میزان محصولات کشاورزی را بهبود بخشیده‌اند. Shahhosseini و همکاران [۱۵] از مدل‌های مختلفی مانند جنگل تصادفی، LightGBM، XGBoost، رگرسیون لاسو (LASSO) و رگرسیون خطی برای پیش‌بینی میزان محصول ذرت استفاده کردند و با طراحی چهار مدل تجمیعی شامل رگرسیون انباشته، لاسو انباشته، جنگل تصادفی انباشته و LightGBM انباشته نشان دادند که تجمیع وزنی بهینه و تجمیع متوسط دقیق‌ترین نتایج را ارائه می‌دهند. همچنین، Sun و همکاران [۱۶] مدل جدیدی مبتنی بر ترکیب شبکه عصبی پیچشی و حافظه طولانی کوتاه‌مدت پیشنهاد کردند که با داده‌های سنجش از دور سری زمانی و خصوصیات خاک، به پیش‌بینی دقیق محصول در سطح شهرستان‌های کمر بند ذرت ایالات متحده کمک کرد.

مطالعات دیگری نیز به ترکیب مدل‌های یادگیری عمیق پرداخته‌اند. Ghazaryan و همکاران [۱۷] با ترکیب مدل‌های شبکه عصبی پیچشی ۳ بعدی و حافظه طولانی کوتاه‌مدت و استفاده از داده‌های MODIS نشان دادند که می‌توان میانگین درصد خطای پیش‌بینی ذرت و سویا را به ترتیب به ۱۰/۳ و ۹/۶ درصد کاهش داد. Khaki و همکاران [۱۸] نیز چارچوب شبکه عصبی پیچشی-شبکه عصبی بازگشتی را معرفی کردند که در مقایسه با روش‌های متداول مانند جنگل تصادفی و شبکه عصبی پیشخور عمیق دقت بالاتری در پیش‌بینی محصول ذرت و سویا در کل کمر بند ذرت ایالات متحده داشت.

تمرکز برخی مطالعات بر بررسی شرایط خاص بوده است. Kim و همکاران [۴] شش مدل هوش مصنوعی را برای پیش‌بینی میزان محصول ذرت در شرایط آب و هوایی شدید بررسی کردند و نشان دادند که مدل شبکه عصبی عمیق با بهینه‌سازی فرایارامترها، دقت پیش‌بینی را در شرایط موج گرما ۳۰-۷۷ درصد بهبود می‌بخشد. این نتایج مکمل یافته‌های Kim و همکاران [۵] است که مدل‌های مختلفی را برای پیش‌بینی میزان محصول ذرت و سویا در بازه زمانی ژوئیه تا اوت بررسی کرده و نشان دادند مدل شبکه عصبی عمیق بهترین عملکرد را با کمترین میانگین خطای مطلق ارائه می‌دهد.

مطالعات اخیر به طور گسترده بر توسعه روش‌های ترکیبی برای بهبود دقت پیش‌بینی محصولات زراعی تمرکز داشته‌اند. سبزیلی یمقانی و همکاران [۱۹] یک رویکرد هیبریدی مبتنی بر الگوریتم XGBoost و بهینه‌ساز عقاب طلایی را برای پیش‌بینی عملکرد سویا با استفاده از داده‌های ماهواره‌ای و هواشناسی معرفی کردند. نتایج این مطالعه نشان داد که مدل ترکیبی GEO-XGBoost در پیش‌بینی عملکرد سویا با ضریب همبستگی ۰/۹۳۷۷ و شاخص خطای ریشه میانگین مربعات ۰/۲۳۹۴ تن در هکتار عملکرد مطلوبی دارد. Kwak و همکاران [۲۰] با بررسی مدل هیبریدی شبکه عصبی پیچشی و جنگل تصادفی، دریافتند که این ترکیب در نقشه‌برداری زودهنگام محصول دقت طبقه‌بندی بالایی ارائه می‌دهد. علاوه بر این، Shahhosseini و همکاران [۲۱] با

دقت پیش‌بینی میزان محصول ذرت افزایش یابد. این رویکرد می‌تواند به بهبود مدیریت کشاورزی و کاهش خسارات ناشی از نوسانات آب‌وهوایی کمک شایانی کند.

نوآوری کلیدی این تحقیق، ادغام الگوریتم XGBoost به‌عنوان یک مدل پیش‌بینی قوی با الگوریتم POA برای بهینه‌سازی موثر پارامترهای مدل است. همچنین، در این پژوهش تأثیر مراحل مختلف فنولوژی گیاه بر دقت پیش‌بینی مورد بررسی قرار گرفته و یک دوره زمانی بهینه برای تخمین میزان محصول انتخاب شده است. این دوره شامل سه بازه زمانی مختلف است: (۱) ماه‌های مه تا سپتامبر (کل فصل رشد)، (۲) ماه‌های ژوئیه تا اوت (دوره تولید عمده)، و (۳) ترکیب‌های مختلف ماه‌ها که بالاترین همبستگی را میان متغیرهای ورودی و میزان محصول نشان می‌دهند. فرض اصلی این تحقیق این است که پیش‌بینی دقیق میزان محصول می‌تواند تنها بر اساس داده‌های مربوط به بازه ژوئیه تا اوت انجام شود. این موضوع امکان پیش‌بینی میزان محصول سال را پیش از زمان برداشت فراهم می‌کند و می‌تواند تصمیم‌گیری‌های بهینه‌تری را در مدیریت زراعت و برنامه‌ریزی کشاورزی میسر سازد.

مواد و روش‌ها

منطقه مورد مطالعه

ایالات متحده به عنوان بزرگ‌ترین صادرکننده غلات در جهان شناخته می‌شود، و این پژوهش بر پنج ایالت غرب میانه آن شامل ایلینوی، آیووا، مینه‌سوتا، داکوتای شمالی و داکوتای جنوبی، که در آن‌ها ذرت محصول غالب است، متمرکز شده است. محدوده مورد بررسی شامل ۴۰۷ شهرستان در این پنج ایالت است (۱۰۲ شهرستان در ایلینوی، ۹۹ شهرستان در آیووا، ۸۷ شهرستان در مینه‌سوتا، ۵۳ شهرستان در داکوتای شمالی و ۶۶ شهرستان در داکوتای جنوبی). این شهرستان‌ها بر اساس معیار مساحت زمین‌های زراعی، که بیش از ۵۰ درصد از سطح هر شهرستان را تشکیل می‌دهد، انتخاب شده‌اند. (شکل ۱). در قسمت پایین شکل ۱ مقادیر تصویر ماهواره‌ای (Leaf Area Index) LAI مودیس به صورت موزاییک شده (تایل‌های h11v04، h10v05، h11v05) برای تاریخ ۹ مه ۲۰۱۱ (روز ۱۲۹ از ابتدای سال) به همراه مرزبندی پنج ایالت و ۴۰۷ شهرستان ارائه شده است.

داده‌های مورد استفاده

ویژگی‌های داده‌های به‌کاررفته در این پژوهش، شامل قدرت تفکیک زمانی و مکانی، نام محصول (در صورت استفاده از داده‌های خاص یک سنجنده) و منبع آن‌ها، در جدول ۱ ارائه شده است. داده‌ها در چهار گروه اصلی طبقه‌بندی شده‌اند: سنجنش از دور (۷ نوع)، آب‌وهوایی (۴ نوع)، زمین زراعی (۱ نوع)، و بازده محصول (۱ نوع). این داده‌ها طی بازه زمانی ۲۰۱۱ تا ۲۰۲۰ و برای فصل رشد، یعنی ماه‌های مه تا سپتامبر، جمع‌آوری شده‌اند. یکی از منابع کلیدی داده‌ها، لایه داده زراعی (CDL) دپارتمان کشاورزی ایالات متحده، مجموعه‌ای از داده‌های رستری

ادغام مدل‌های یادگیری ماشین و متغیرهای مرتبط با رطوبت خاک، توانستند دقت پیش‌بینی عملکرد محصولات را در کمربند ذرت ایالات متحده بهبود بخشند. همچنین، مطالعه دیگری از Shahhosseini و همکاران [۲۲] نشان داد که مدل‌های تجمیعی، شامل شبکه عصبی پیچشی و شبکه عصبی عمیق، عملکرد بهتری نسبت به مدل‌های منفرد مانند جنگل تصادفی، XGBoost و LightGBM در پیش‌بینی بازده محصولات دارند. این نتایج بر اهمیت استفاده از رویکردهای ترکیبی در افزایش دقت و کارایی پیش‌بینی عملکرد محصولات کشاورزی تأکید می‌کند.

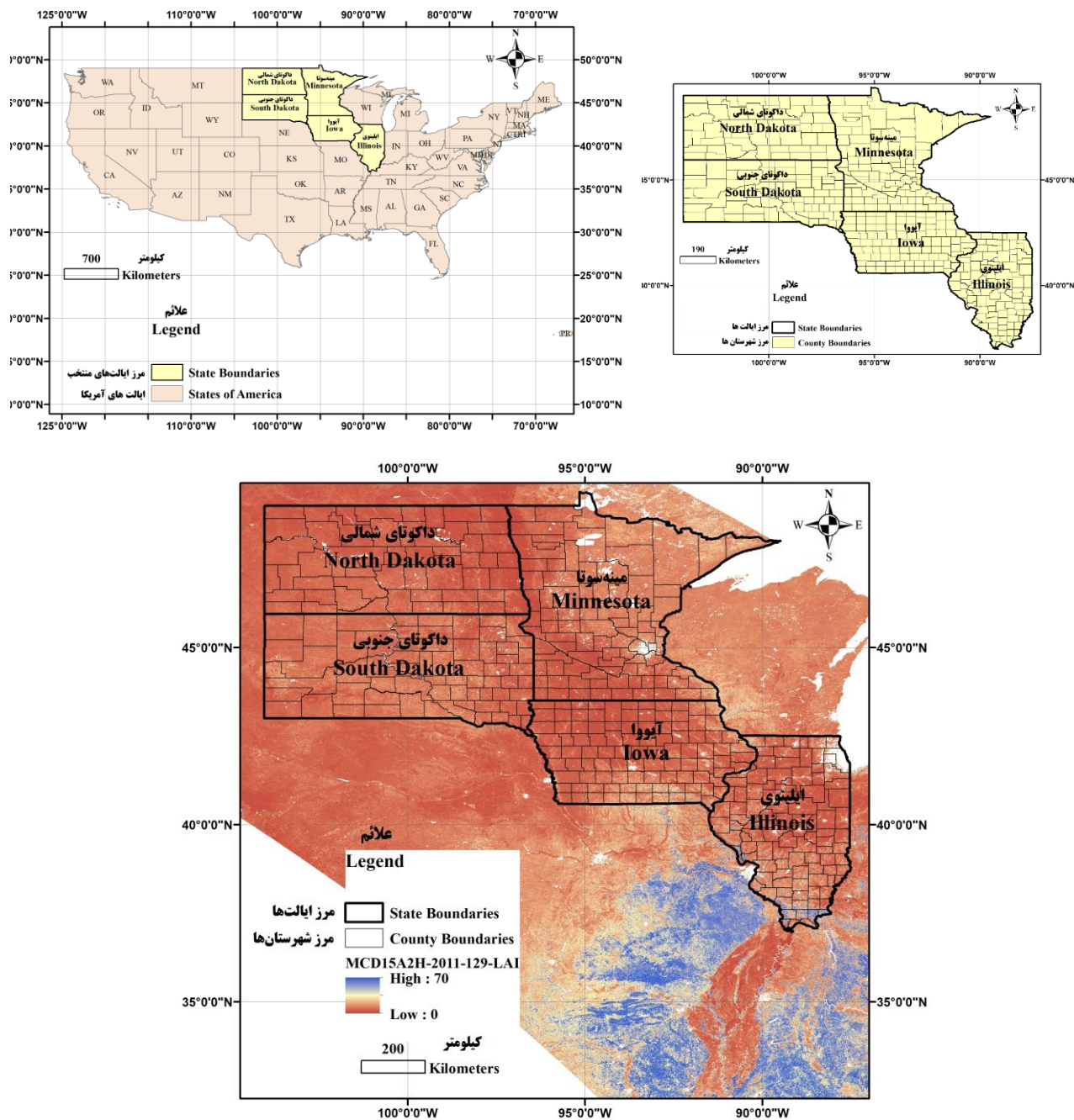
تحقیقات اخیر به استفاده از تکنیک‌های نوآورانه و داده‌های متنوع متمرکز بوده‌اند. Colombo-Mendoza و همکاران [۲۳] یک سیستم کشاورزی هوشمند مبتنی بر داده‌کاوی اینترنت اشیا برای پیش‌بینی محصولات طراحی کردند. Schmidt و همکاران [۲۴] از شبکه‌های عصبی مصنوعی برای کاهش ریسک پایه در بیمه‌های کشاورزی استفاده کردند. van Klompenburg و همکاران [۲۵] نیز مرور جامعی از مطالعات پیش‌بینی میزان محصول ارائه دادند که بر کاربرد مدل‌هایی چون رگرسیون خطی، ماشین بردار پشتیبان و شبکه‌های عصبی تأکید داشت. Meghraoui و همکاران [۲۶] نشان دادند که معماری‌های شبکه عصبی پیچشی و حافظه طولانی کوتاه‌مدت در پیش‌بینی محصولات کشاورزی غالب هستند. علاوه بر این، Oikonomidis و همکاران [۲۷] با بررسی مدل‌های هیبریدی مختلف، به برتری مدل شبکه عصبی پیچشی-شبکه عصبی عمیق در دقت پیش‌بینی محصولات اشاره کردند. این مطالعات نشان می‌دهند که استفاده از داده‌های متنوع نظیر تصاویر ماهواره‌ای، داده‌های هواشناسی و خصوصیات خاک، همراه با مدل‌های پیشرفته یادگیری عمیق و تجمیعی، دقت پیش‌بینی محصولات کشاورزی را بهبود می‌بخشد. توسعه مدل‌های ترکیبی جدید می‌تواند به کشاورزان و مدیران در مدیریت منابع و افزایش بهره‌وری کمک کند.

بر اساس بررسی پیشینه تحقیقات، می‌توان نتیجه گرفت که تکنیک‌های یادگیری ماشین و سنجنش از دور در سال‌های اخیر پیشرفت‌های چشمگیری داشته‌اند و به ابزارهای کلیدی در پیش‌بینی میزان تولیدات کشاورزی تبدیل شده‌اند. پژوهش‌ها حاکی از آن است که بهره‌گیری از داده‌های متنوع، از جمله تصاویر ماهواره‌ای، اطلاعات هواشناسی، و ویژگی‌های خاک، در کنار به‌کارگیری روش‌های پیشرفته یادگیری عمیق و الگوریتم‌های گروهی، توانسته است دقت پیش‌بینی‌ها را به میزان قابل توجهی بهبود بخشد. از این‌رو، انجام تحقیقات بیشتر در این حوزه و توسعه مدل‌های نوآورانه ترکیبی می‌تواند نقش مؤثری در کمک به کشاورزان و مدیران کشاورزی برای بهینه‌سازی فرآیند تولید و مدیریت منابع ایفا کند.

هدف اصلی این پژوهش توسعه یک رویکرد ترکیبی است که از الگوریتم XGBoost برای پیش‌بینی و الگوریتم بهینه‌سازی پلیکان (POA) برای تنظیم پارامترهای مدل استفاده می‌کند. در این مطالعه، تلاش شده است تا با بهره‌گیری از داده‌های گسترده و متنوع ماهواره‌ای و هواشناسی،

محاسباتی را کنترل کرده است. همچنین، به علت تلفیق انواع داده‌های ماهواره‌ای با قدرت تفکیک‌های مختلف (نظیر ۲۵۰ و ۵۰۰ متر)، نتایج نهایی با روش تحلیل ناحیه‌ای (Zonal Analysis) به سطح شهرستان خلاصه می‌شوند. در مقیاس‌های خردتر، می‌توان از داده‌های با تفکیک بالاتر استفاده کرد، اما در این مطالعه سطح شهرستان هدف اصلی است و این قدرت تفکیک، پاسخ‌گوی نیاز پژوهش خواهد بود.

جغرافیایی عمومی است که طبقه‌بندی پوشش زمین را با تأکید بر نقشه‌برداری مزارع و به‌روزرسانی چشم‌انداز کشاورزی در ایالات متحده ارائه می‌دهد. در این پژوهش، داده‌های آب و هوا ۴ کیلومتری PRISM به دلیل همخوانی با سطح تحلیل شهرستانی، دسترسی آسان و یکپارچه، و پشتوانه علمی معتبر انتخاب شده‌اند. تفکیک ۴ کیلومتر ضمن تأمین دقت کافی برای ارزیابی عملکرد محصول در سطح شهرستان، هزینه



شکل ۱: محدوده مورد مطالعه

Fig. 1: The study area

جدول ۱: داده‌های مورد استفاده در تحقیق

Table 1: Data used in the study

منبع Source	نام محصول Product name	قدرت تفکیک Resolution		داده Data	
		زمانی Time	مکانی Place		
NASA EOSDIS	MOD13Q1 MODIS	۱۶ روز / 16 days	۲۵۰ متر / 250 meters	NDVI (Normalized Difference Vegetation Index)	سنجش از دور Remote sensing
	MOD13Q1 MODIS	۱۶ روز / 16 days	۲۵۰ متر / 250 meters	EVI (Enhanced Vegetation Index)	
	MCD15A2H MODIS	۸ روز / 8 days	۵۰۰ متر / 500 meters	LAI (Leaf Area Index)	
	MCD15A2H MODIS	۸ روز / 8 days	۵۰۰ متر / 500 meters	FPAR (Fraction of Photosynthetically Active Radiation)	
	MYD17A2H MODIS	۸ روز / 8 days	۵۰۰ متر / 500 meters	GPP (Gross Primary Production)	
	MOD16A2 MODIS	۸ روز / 8 days	۵۰۰ متر / 500 meters	ET (Evapotranspiration)	
NASA GES DISC	GLDAS_NOAH025_M	ماهانه / Monthly	۲۵ درجه / 25°	SM (Soil Moisture)	آب و هوا Meteorological
گروه آب و هوای PRISM	-	ماهانه / Monthly	۴ کیلومتر / 4 km	PPT (Precipitation)	
		ماهانه / Monthly	۴ کیلومتر / 4 km	Tmax (Maximum Temperature)	
		ماهانه / Monthly	۴ کیلومتر / 4 km	Tmin (Minimum Temperature)	
		ماهانه / Monthly	۴ کیلومتر / 4 km	Tmean (Average Temperature)	
دپارتمان کشاورزی ایالات متحده United States Department of Agriculture	-	سالانه / Annual	۳۰ متر / 30 meters	Cropland	زمین زراعی Cropland
دپارتمان کشاورزی ایالات متحده United States Department of Agriculture	-	سالانه / Annual	شهرستان / County	ذرت / Corn	بازده محصول Crop Yield

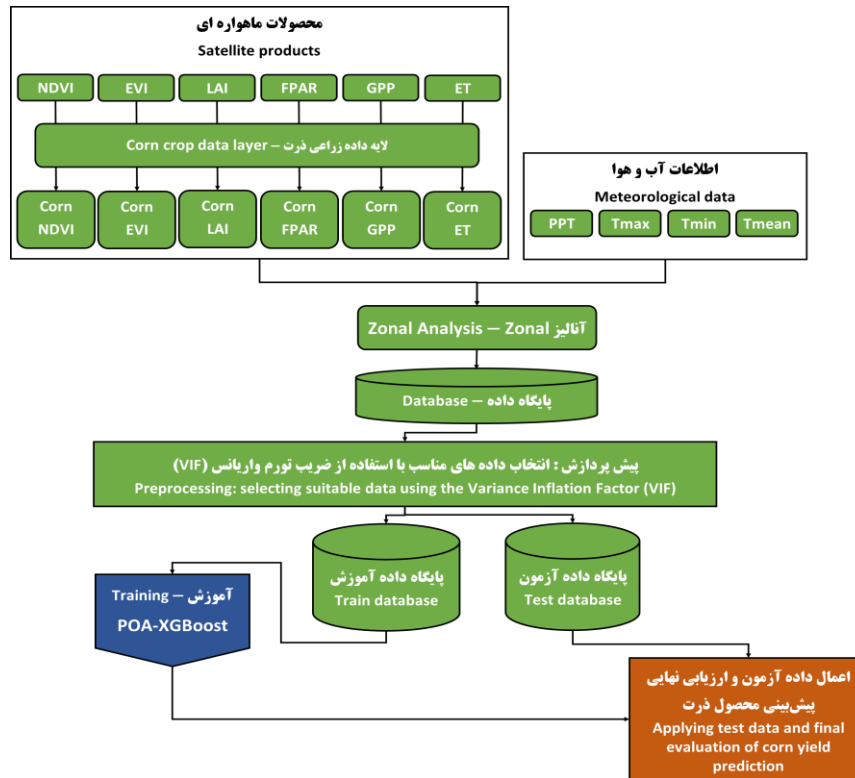
روش‌شناسی تحقیق

بررسی تحقیقات پیشین نشان داد که مجموعه داده‌های مشابه جدول ۱ در پیش‌بینی میزان محصول ذرت مورد استفاده قرار گرفته‌اند. در این پژوهش نیز از داده‌های جدول ۱ بهره‌گیری شده و با تمرکز بر الگوریتم یادگیری گروهی XGBoost، تلاش می‌شود تا مدلی بهبودیافته ارائه شود. برای این منظور، بهینه‌سازی پارامترهای الگوریتم XGBoost با استفاده از الگوریتم POA انجام می‌شود. شمای کلی فرآیند تحقیق در شکل ۲ ارائه شده است. در ابتدا، داده‌های سنجش از دور با استفاده از لایه داده‌های زراعی تفکیک شده و نواحی فاقد محصولات زراعی مدنظر در تصاویر به‌عنوان مقادیر ناموجود در نظر گرفته می‌شوند. این فرآیند برای محصول ذرت اعمال می‌شود و تنها پیکسل‌های مرتبط با این محصول دارای مقدار خواهند بود. سپس، آنالیز Zonal بر اساس سطح شهرستان برای تمامی داده‌ها انجام شده و داده‌های مجزا برای محصولات زراعی ذرت تولید می‌شود. در گام بعد، شاخص آماری VFI برای انتخاب داده‌های مناسب مدل داده‌کاوی مورد استفاده قرار می‌گیرد و داده‌های منتخب وارد فرآیند داده‌کاوی می‌شوند. الگوریتم XGBoost با استفاده از داده‌های ۹ سال برای آموزش و یک سال برای آزمون، طی بازه زمانی ۲۰۱۱ تا ۲۰۲۰، آموزش داده می‌شود. تعداد داده‌های مورد استفاده در این تحقیق به طور کلی ۳۳۷۱ مورد برای ۱۰ سال می‌باشد. برای سال‌های ۲۰۱۱ تا ۲۰۲۰ به ترتیب ۳۷۰، ۳۶۱، ۳۴۷، ۳۴۹، ۳۳۳، ۳۵۶، ۳۴۱، ۲۹۰، ۲۶۴ و ۳۶۰ مورد به عنوان داده‌های آزمون انتخاب شدند. میانگین شاخص‌های ارزیابی به‌عنوان دقت نهایی گزارش می‌شود و در

تمامی مراحل، بهینه‌سازی فرایپارامترهای XGBoost با الگوریتم POA صورت می‌گیرد.

الگوریتم XGBoost

الگوریتم XGBoost یکی از پرکاربردترین و قدرتمندترین روش‌های مبتنی بر تقویت گرادیان (Gradient Boosting) است که برای حل مسائل طبقه‌بندی و رگرسیون به کار می‌رود. این روش با ترکیب چندین درخت تصمیم (Decision Trees)، مدلی قوی و پایدار ارائه می‌دهد که در هر مرحله خطای به‌جامانده از مرحله قبل را اصلاح می‌کند. در ابتدا، گرادیان و هسیان (مشتق‌های مرتبه اول و دوم تابع هزینه) برای داده‌ها محاسبه می‌شوند تا روند کاهش خطا هدفمندتر شود. سپس یک درخت تصمیم جدید ساخته می‌شود که انشعاب‌های آن (Splits) براساس مقادیر گرادیان و هسیان انتخاب می‌گردند؛ این امر باعث می‌شود پیش‌بینی‌های نادرست مراحل قبل تصحیح شود و دقت مدل به طور تدریجی افزایش یابد. فرآیند تا زمان همگرایی یا رسیدن به تعداد معینی درخت تکرار می‌شود و در هر مرحله، برای جلوگیری از برازش بیش از حد (Overfitting)، از مکانیزم‌های منظم‌سازی (Regularization) و هرس (Pruning) استفاده می‌شود. افزون بر این، به‌واسطه پیاده‌سازی موازی و ساختار بهینه پردازشی، الگوریتم XGBoost نسبت به روش‌های مشابه، سرعت بالایی در آموزش دارد و در بسیاری از رقابت‌های علمی و کاربردهای تجاری به یک ابزار کلیدی تبدیل شده است. شبه‌کد الگوریتم XGBoost برای درک بهتر فرآیند در الگوریتم ۱ آورده شده است [۲۸].



شکل ۲: شمای کلی تحقیق
Fig. 2: General overview of the study

خود را بهینه تر می کنند. این فرآیند به طور مداوم تکرار می شود تا الگوریتم به بهترین جواب ممکن دست یابد. مزیت اصلی POA نسبت به سایر روش های بهینه سازی، تعادل مناسب بین جستجو و بهره برداری، سرعت همگرایی بالا، و توانایی فرار از بهینه های محلی (Local Optima) است که باعث می شود در بسیاری از مسائل مهندسی، پردازش تصویر، یادگیری ماشین، و بهینه سازی توابع ریاضی عملکرد موفق داشته باشد. جزئیات مربوط به این الگوریتم در قالب شبه کد در الگوریتم ۲ ارائه شده است [۲۹]

Algorithm 2. Pseudo-code of POA.

Start POA.

Input the optimization problem information.

Determine the POA population size (N) and the number of iterations (T).

Initialization of the position of pelicans and calculate the objective function.

For t = 1:T

Generate the position of the prey at random.

For l = 1:N

Phase 1: Moving towards prey (exploration phase).

For j = 1:m

Calculate new status of the jth dimension.

End.

Update the ith population member.

Phase 2: Winging on the water surface (exploitation phase).

For j = 1:m.

Calculate new status of the jth dimension.

End.

Update the ith population member.

Algorithm 1. Pseudo-code of XGBoost

Input: instance set of current nodes

Input: feature dimension

$J(p) = 0$

$G = \sum_{i \in I} g_i \quad H = \sum_{i \in I} h_i$

For k = 1 to n do

$G_L = 0 \quad H_L = 0$

For j in sorted do

$G_L = G_L + g_j \quad H_L = H_L + h_j$

$G_R = G - G_L \quad H_R = H - H_L$

$score = \max(score, J(p))$

End

End

Output: Split with max score

//الگوریتم POA

الگوریتم POA یک روش فراابتکاری هوشمند است که بر اساس رفتار پلیکان ها هنگام شکار طراحی شده و برای حل مسائل بهینه سازی مورد استفاده قرار می گیرد. ایده اصلی این روش از دو رفتار کلیدی پلیکان ها در طبیعت الهام گرفته شده است: شیرجه زدن به سمت طعمه که نشان دهنده فاز جستجو (Exploration) است و بال زدن روی سطح آب که نشان دهنده فاز بهره برداری (Exploitation) است. در فاز اول، پلیکان ها به صورت تصادفی در فضای جستجو حرکت می کنند تا گزینه های مناسب برای یافتن جواب بهینه را شناسایی کنند. سپس، در فاز دوم، روی بهترین موقعیت های شناسایی شده تمرکز کرده و موقعیت

ذرت ایجاد شد. به منظور بررسی وجود هم خطی بین ۱۱ متغیر توضیحی، آزمون چندخطی با استفاده از ضریب تورم واریانس (Variation Inflation Factor: VIF) انجام گرفت تا از تأثیرات احتمالی هم خطی بر دقت مدل جلوگیری شود. VIF برای ۱۱ متغیر توضیحی x_i به صورت زیر بیان می شود:

$$VIF(x_i) = \frac{1}{1 - R_i^2} \quad (۵)$$

در اینجا R_i^2 نشان دهنده ضریب تعیین معادله رگرسیون است، که در آن x_i به عنوان متغیر وابسته و سایر متغیرهای x به عنوان متغیرهای مستقل در نظر گرفته می شوند. برای محصولات ذرت، ابتدا مقدار شاخص VIF برای هر ۱۱ متغیر اولیه محاسبه شد. با توجه به اینکه مقدار برخی از این متغیرها از حد آستانه ۵ فراتر رفت [۵]، به منظور کاهش مقدار VIF، متغیرهایی با بالاترین مقدار VIF به صورت تدریجی حذف شدند و مقادیر VIF مجدداً محاسبه گردید. در این فرآیند، متغیرهای T_{MAX} ، ET ، $NDVI$ ، T_{MEAN} و $FPAR$ به ترتیب با مقادیر VFI برابر با ۷۲۳۴۲۵۷۰، ۳۳/۹۹، ۲۶/۰۴، ۱۴/۴۸ و ۶/۴۵ حذف شدند. این روند حذف تدریجی ادامه یافت تا مقدار شاخص VIF برای تمامی متغیرهای باقی مانده به مقادیر قابل قبول با حد آستانه ۵ کاهش یابد.

جدول ۲: مقدار VFI متغیرهای منتخب (مه تا سپتامبر) در برابر میزان محصول ذرت، ۲۰۱۱-۲۰۲۰

داده ها	ضریب تورم واریانس
Data	VIF
NDVI	-
EVI	3.9465
LAI	2.9043
FPAR	-
GPP	4.3388
ET	-
SM	2.6517
PPT	1.2709
T_{MAX}	-
T_{MIN}	1.7678
T_{MEAN}	-

اثر فنولوژی

به طور کلی، کاشت ذرت معمولاً بین اواخر آوریل تا اواسط مه انجام می شود، این محصول در ماه سپتامبر به بلوغ می رسد و تا اواخر اکتبر برداشت می شود. برای بررسی اثرات فنولوژی بر تولید محصول، ۱۳ ترکیب زمانی مختلف از داده ها استخراج شد. این ترکیبات شامل یک دوره نماینده برای کل فصل رشد (GS) از مه تا سپتامبر، پنج دوره برای ماه های مجزا، چهار ترکیب دوماهه متوالی، و سه ترکیب سه ماهه متوالی بودند (جدول ۳). با تحلیل ضرایب همبستگی میان متغیرهای ورودی و میزان محصول، سه دوره متمایز شناسایی شدند: (۱) کل فصل رشد (GS) از زمان کاشت تا برداشت، (۲) دوره اصلی تولید ذرت از ژوئیه تا اوت (JA)، و (۳) ترکیب بهینه ماه هایی که بیشترین همبستگی را بین متغیرهای ورودی و میزان محصول نشان دادند. ترکیب بهینه، که برای

End.

Update best candidate solution.

End.

Output best candidate solution obtained by POA.

End POA.

ارزیابی دقت

برای ارزیابی صحت و دقت روش های پیش بینی میزان محصول، پنج شاخص کلیدی شامل میانگین درصد مطلق خطا (MAPE)، میانگین انحراف خطا (MBE)، میانگین خطای مطلق (MAE)، ریشه میانگین مربعات خطا (RMSE) و ضریب تعیین (R^2) در این پژوهش به کار گرفته شده اند. شاخص RMSE معمولاً میزان پراکندگی نتایج پیش بینی را نشان می دهد، در حالی که شاخص های MAPE و MAE برای ارزیابی میزان انحراف نتایج پیش بینی از مقادیر واقعی استفاده می شوند. این شاخص ها نقش اساسی در تحلیل عملکرد مدل های پیش بینی دارند و به صورت زیر تعریف می شوند [۳۰-۳۲]:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \quad (۱)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (۲)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (۳)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (۴)$$

در این تعریف ها $\hat{y}_i$ نشان دهنده مقدار پیش بینی شده، y_i نشان دهنده مقدار واقعی و n تعداد کل نمونه ها است. هر یک از این شاخص ها به صورت خاص برای ارزیابی جنبه های مختلف عملکرد مدل مورد استفاده قرار می گیرند. در نهایت، ضریب تعیین (R^2) به عنوان یک متریک آماری عمل می کند که نشان می دهد چه مقدار از تغییرات در یک متغیر وابسته می تواند توسط متغیر (های) مستقل در یک مدل رگرسیونی توضیح داده شود. مقدار R^2 بین ۰ و ۱ متغیر است، که مقدار بالاتر آن نشان دهنده تطابق بهتر مدل با داده ها است.

نتایج پژوهش

پیش پردازش داده ها

در این پژوهش، ابتدا پیکسل های زمین زراعی که در CDL به عنوان ذرت (با شناسه ID=1) ثبت شده بودند، استخراج شدند. سپس، توزیع متغیرهای ورودی شامل NDVI، EVI، LAI، FPAR، ET، GPP، PPT، TMAX، TMIN و TMEAN از شبکه CDL استخراج شد. برای تطبیق داده ها با آمار محصول در سطح شهرستان، تمامی متغیرها با استفاده از یک عملیات میانگین گیری منطقه ای، مطابق با مرزهای چندضلعی هر شهرستان، تجمیع شدند. در نهایت، پایگاه داده ای مختص مناطق کشت

در این بخش، نتایج توسعه مدل ترکیبی XGBoost و POA برای پیش‌بینی میزان محصول ذرت با بهره‌گیری از داده‌های ماهواره‌ای و هواشناسی ارائه شده است. برای بهبود عملکرد مدل، پارامترهای مربوط به بهینه‌سازی الگوریتم XGBoost و تنظیمات الگوریتم POA مورد بررسی قرار گرفتند. جزئیات این پارامترها و مقادیر مربوطه در جدول ۴ ارائه شده است.

در این پژوهش، با توجه به اینکه دوره JA (ژوئیه تا اوت) دوره اصلی تولید ذرت محسوب می‌شود، داده‌های این بازه زمانی برای پیش‌بینی میزان محصول مورد استفاده قرار گرفت. انتخاب این دوره امکان پیش‌بینی میزان محصول را پیش از برداشت فراهم کرده و به تصمیم‌گیری‌های زراعی کمک می‌کند. جدول ۵ نتایج الگوریتم ترکیبی POA-XGBoost را برای محصولات ذرت در دوره JA طی سال‌های ۲۰۱۱ تا ۲۰۲۰ نشان می‌دهد. این نتایج عملکرد مدل در پیش‌بینی میزان محصول را با شاخص‌های ارزیابی مختلف مانند MAE، MBE، RMSE، و MAPE و ضریب همبستگی R نمایش می‌دهد.

پیش‌بینی میزان محصول ذرت مطلوب‌ترین است، در جدول ۳ با رنگ کرم مشخص شده است.

به طور کلی، ذرت بین اواخر آوریل و اواسط مه کاشته می‌شود، در سپتامبر بالغ می‌شود و تا اواخر اکتبر برداشت می‌شود. برای تجزیه و تحلیل اثرات فنولوژی، ۱۳ مورد با استفاده از ترکیب‌های مختلف ماه استخراج شد، از جمله یک مورد نماینده کل فصل رشد (GS) بین ماه‌های مه و سپتامبر، پنج مورد از ماه‌های جداگانه، چهار مورد از دو ماه متوالی، و سه مورد مربوط به سه ماه متوالی (جدول ۳). با در نظر گرفتن ضرایب همبستگی بین متغیرهای ورودی و میزان محصول، سه دوره متمایز شد: (۱) کل GS از کاشت تا برداشت، (۲) ژوئیه تا اوت (JA) که دوره اصلی تولید ذرت است و (۳) ترکیب بهینه ماه‌ها با همبستگی بالا بین متغیرهای نهاده و میزان محصول. ترکیب بهینه برای میزان محصول ذرت بر اساس ترکیب داده‌های ماه‌های مختلف در جدول ۳ با رنگ کرم مشخص شده است.

توسعه مدل پیش‌بینی میزان محصول ذرت

جدول ۳: ضرایب همبستگی بین متغیرهای انتخاب شده و میزان محصول ذرت با توجه به ترکیب ماه‌های ۲۰۲۰-۲۰۱۱

Table 3: Correlation coefficients between selected variables and corn yield considering monthly combinations, 2011-2020

GS	JAS	JJA	MJJ	AS	JA	JJ	MJ	Sep	Aug	Jul	Jun	May	
0.788	0.778	0.864	0.647	0.839	0.839	0.739	0.349	0.228	0.706	0.830	0.481	-0.015	EVI
0.521	0.626	0.580	0.290	0.651	0.652	0.372	-0.113	0.147	0.617	0.530	-0.008	-0.261	LAI
0.698	0.749	0.754	0.542	0.761	0.762	0.638	0.074	0.392	0.705	0.706	0.190	-0.087	GPP
0.501	0.436	0.449	0.425	0.421	0.421	0.369	0.308	0.266	0.334	0.320	0.218	0.236	PPT
0.383	0.360	0.314	0.301	0.209	0.209	0.284	0.378	0.472	0.323	0.057	0.459	0.293	T _{MIN}
0.556	0.578	0.550	0.513	0.562	0.563	0.536	0.463	0.550	0.536	0.565	0.479	0.423	SM

جدول ۴: مقادیر پارامترهای قابل تنظیم در الگوریتم POA-XGBoost

Table 4: Tunable parameter values in the poa-xgboost algorithm

Usage / استفاده	Values / مقادیر مربوطه	Adjustable parameter / پارامتر قابل تنظیم
الگوریتم XGBoost	20-400	n_estimators
الگوریتم XGBoost	3-18	max_depth
الگوریتم XGBoost	0-9	gamma
الگوریتم XGBoost	0-1	learning_rate
الگوریتم POA	30	جمعیت اولیه
الگوریتم POA	10	تعداد اجرا

جدول ۵: خلاصه آمار اعتبارسنجی از ۱۰ دور آزمایش برای میزان محصول ذرت در ژوئیه تا اوت (دوره تولید عمده) برای ۲۰۲۰-۲۰۱۱

Table 5: Summary statistics of validation from 10 rounds of testing for corn yield during July to August (major production period), 2011-2020

R	MAPE	RMSE	MAE	MBE	تعداد نمونه‌ها Number of samples	
0.8499	0.1004	1.1010	0.8933	0.1242	370	2011
0.8865	0.1995	1.2944	0.9704	-0.0319	361	2012
0.9067	0.0704	0.8444	0.6489	0.0067	347	2013
0.9372	0.0626	0.8102	0.6159	0.0066	349	2014
0.9097	0.0631	0.8841	0.6561	0.0862	333	2015
0.8984	0.0712	1.0081	0.7760	0.2176	356	2016
0.9071	0.0870	1.1113	0.9355	0.4457	341	2017
0.8083	0.0837	1.3113	1.0340	0.8039	290	2018

R	MAPE	RMSE	MAE	MBE	تعداد نمونه‌ها Number of samples	
0.8205	0.0790	1.0653	0.8785	0.1667	264	2019
0.9237	0.0657	0.8372	0.6434	-0.0720	360	2020
0.8848	0.0883	1.0267	0.8052	0.1754	میانگین Average	

بلکه با بررسی اثرات تغییرات آب‌وهوا، به بهبود مدیریت محصولات و امنیت غذایی کمک می‌کنند. پیشرفت‌های اخیر در تکنیک‌های یادگیری ماشین و دسترسی به داده‌های بزرگ، نظیر داده‌های سنجش از دور و داده‌های عملکرد تاریخی محصولات، تحولی در تجزیه و تحلیل کشاورزی ایجاد کرده است. این فناوری‌ها با تبدیل داده‌های پیچیده به بینش‌های کاربردی، امکان بهینه‌سازی تولید، تخصیص منابع و نظارت بر عملکرد محصولات در زمان واقعی را فراهم می‌کنند و راهکاری نوین برای مقابله با چالش‌های کشاورزی مدرن ارائه می‌دهند [۷، ۱۰، ۱۲].

در این پژوهش، یک چارچوب نوین برای پیش‌بینی میزان محصول ذرت در سطح شهرستان‌های پنج ایالت غرب میانه ایالات متحده، با استفاده از ترکیب الگوریتم‌های XGBoost و POA پیشنهاد شد. نتایج نشان داد که پیش‌بینی‌های دقیق می‌توانند با بهره‌گیری از داده‌های دوره ژوئیه تا اوت (دوره تولید عمده) و پیش از زمان برداشت انجام شوند. عملکرد چارچوب پیشنهادی این مطالعه با میانگین ضریب همبستگی ۱۰ ساله برابر ۰/۸۸۴۸ نشان‌دهنده دقت مناسب در پیش‌بینی‌ها بود. بررسی سال به سال نتایج نشان داد که دقت پیش‌بینی‌ها در سال‌های ۲۰۱۳، ۲۰۱۴، و ۲۰۲۰ به اوج خود رسیده و مقادیر خطا مانند MAE و RMSE در این سال‌ها کمترین مقدار را داشتند، در حالی که در سال‌های ۲۰۱۷ و ۲۰۱۸ دقت کاهش یافته و با یاس مثبت پیش‌بینی‌ها به بیشترین مقدار رسید. با این حال، در سال‌های پایانی، کاهش خطاها و بهبود دقت کلی، اهمیت چارچوب پیشنهادی در ارائه پیش‌بینی‌های دقیق برای کشاورزان و برنامه‌ریزان را برجسته می‌کند. همچنین، استفاده از تصاویر ماهواره‌ای با شاخص وضوح بالا می‌تواند دقت این پیش‌بینی‌ها را بیش از پیش افزایش دهد.

مشارکت نویسندگان

در این مقاله نویسندگان به نسبت برابر مشارکت داشته‌اند.

تشکر و قدردانی

بدین وسیله از کلیه عزیزانی که ما را در انجام این تحقیق یاری کردند، صمیمانه تشکر می‌کنیم.

تعارض منافع

«هیچ‌گونه تعارض منافع توسط نویسندگان بیان نشده است.»

منابع و مأخذ

[1] Sierra-Forero B-L, Baron-Velandia J, Vanegas-Ayala S-C. Assessment of the relevance of features associated with corn

در طول دوره ۲۰۱۱ تا ۲۰۲۰، روند تغییرات مقادیر خطای پیش‌بینی نشان‌دهنده الگوهای متفاوتی از دقت و بایاس در نتایج اعتبارسنجی است. در سال ۲۰۱۱، پیش‌بینی‌ها با دقت مناسبی همراه بودند، به‌طوری که MAPE برابر ۱/۰۴٪ و همبستگی (R) برابر ۰/۸۴۹۹ بود. با این حال، در سال ۲۰۱۲، دقت پیش‌بینی کاهش یافت و مقادیر خطاها مانند MAE و RMSE افزایش یافتند. سال‌های ۲۰۱۳ و ۲۰۱۴ با بهبود قابل توجهی همراه بودند؛ به‌طوری که MAPE به ترتیب به ۷/۰۴٪ و ۶/۲۶٪ کاهش یافت و مقدار همبستگی به ۰/۹۳۷۲ در سال ۲۰۱۴ رسید که بالاترین میزان همبستگی در طول دوره بود.

در سال‌های ۲۰۱۵ تا ۲۰۱۸، روند پیش‌بینی‌ها دستخوش تغییراتی شد. در سال ۲۰۱۵، مقادیر خطاها اندکی افزایش یافت، اما همچنان در محدوده قابل قبول باقی ماندند. سال ۲۰۱۶ با افزایش بایاس مثبت و مقادیر خطاها همراه بود، اما همچنان همبستگی بالایی بین مقادیر پیش‌بینی‌شده و واقعی مشاهده شد. سال ۲۰۱۷ با بدترین شرایط بایاس مثبت مواجه شد (MBE برابر ۰/۴۴۵۷)، و در سال ۲۰۱۸ این مقدار به اوج خود یعنی ۰/۸۰۳۹ رسید. کاهش تعداد نمونه‌ها در این سال‌ها ممکن است در افت دقت و افزایش خطاها تأثیرگذار بوده باشد. در این دوره، مقدار R نیز به طور قابل توجهی کاهش یافت و در سال ۲۰۱۸ به پایین‌ترین مقدار خود یعنی ۰/۸۰۸۳ رسید.

در پایان دوره، نتایج سال‌های ۲۰۱۹ و ۲۰۲۰ بهبود چشمگیری نشان دادند. مقدار MBE در سال ۲۰۱۹ به ۰/۱۶۶۷ کاهش یافت و خطاهای پیش‌بینی (MAE و RMSE) نیز روند کاهشی داشتند. سال ۲۰۲۰ با دقت بالا همراه بود؛ به‌طوری که MAPE به ۶/۵۷٪ کاهش یافت و مقدار R به ۰/۹۲۳۷ رسید. به طور کلی، میانگین مقادیر کل دوره نشان‌دهنده روندی نسبتاً پایدار با بهبودهای مقطعی در سال‌های پایانی است. این نتایج نشان می‌دهد که علی‌رغم نوساناتی در مقاطع مختلف، دقت پیش‌بینی‌ها در کل دوره در سطح قابل قبولی قرار داشته است.

نتیجه‌گیری

کشاورزی به‌عنوان یکی از بنیادی‌ترین فعالیت‌های بشری، نقشی حیاتی در تأمین امنیت غذایی و توسعه پایدار ایفا می‌کند. در این میان، فناوری سنجش از دور با ارائه نظارت جامع و دقیق بر رشد محصولات کشاورزی، داده‌هایی عینی و به‌موقع برای برآورد تولید و مدیریت منابع ارائه می‌دهد. پیش‌بینی عملکرد محصولات کشاورزی، مانند ذرت در ایالات متحده، به‌طور خاص در تجارت جهانی غذا و سیاست‌گذاری‌های امنیت غذایی اهمیت ویژه‌ای دارد. این پیش‌بینی‌ها نه تنها مبنایی برای تصمیم‌گیری‌های ملی و بین‌المللی در حوزه کشاورزی پایدار هستند،

- [12] Li Y, Zeng H, Zhang M, Wu B, Zhao Y, Yao X, et al. A county-level soybean yield prediction framework coupled with XGBoost and multidimensional feature engineering. *International Journal of Applied Earth Observation and Geoinformation*. 2023;118:103269. <https://doi.org/10.1016/j.jag.2023.103269>
- [13] Muruganantham P, Wibowo S, Grandhi S, Samrat NH, Islam N. A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sensing*. 2022;14(9):1990. <https://doi.org/10.3390/rs14091990>
- [14] Kim N, Lee Y-W. Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data: A Case of Iowa State: A Case of Iowa State. *Journal of the Korean Society of Surveying Geodesy Photogrammetry and Cartography*. 2016;34(4):383-90. <https://doi.org/10.7848/ksgpc.2016.34.4.383>
- [15] Shahhosseini M, Hu G, Archontoulis SV. Forecasting corn yield with machine learning ensembles. *Frontiers in Plant Science*. 2020; 11: 1120. <https://doi.org/10.3389/fpls.2020.01120>
- [16] Sun J, Lai Z, Di L, Sun Z, Tao J, Shen Y. Multilevel deep learning network for county-level corn yield estimation in the us corn belt. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2020;13:5048-60. <https://doi.org/10.1109/JSTARS.2020.3019046>
- [17] Ghazaryan G, Skakun S, König S, Rezaei EE, Siebert S, Dubovyk O, editors. Crop yield estimation using multi-source satellite image series and deep learning. *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*; 2020: IEEE.
- [18] Khaki S, Wang L, Archontoulis SV. A CNN-RNN framework for crop yield prediction. *Frontiers in Plant Science*. 2020;10:1750. <https://doi.org/10.3389/fpls.2019.01750>
- [19] Sabzali Yameqani A, Alesheikh AA, Majidi M. Development of an Ensemble Learning Approach for Soybean Yield Prediction using Satellite and Meteorological Data. *Journal of Geomatics Science and Technology*. 2024;14(1):1-12. [In Persian] <https://doi.org/10.61186/jgst.14.1.1>
- [20] Kwak G-H, Park C-w, Lee K-d, Na S-i, Ahn H-y, Park N-W. Potential of hybrid CNN-RF model for early crop mapping with limited input data. *Remote Sensing*. 2021;13(9):1629. <https://doi.org/10.3390/rs13091629>
- [21] Shahhosseini M, Hu G, Huber I, Archontoulis SV. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific reports*. 2021a;11(1):1606. <https://doi.org/10.1038/s41598-020-80820-1>
- [22] Shahhosseini M, Hu G, Khaki S, Archontoulis SV. Corn yield prediction with ensemble CNN-DNN. *Frontiers in plant science*. 2021b;12:709008. <https://doi.org/10.3389/fpls.2021.709008>
- [23] Colombo-Mendoza LO, Paredes-Valverde MA, Salas-Zárate MdP, Valencia-García R. Internet of Things-driven data mining crop yield prediction in Colombia, a country in the Neotropical zone. *International Journal of Information Technology*. 2024;16(4):2129-38. <https://doi.org/10.1007/s41870-024-01762-9>
- [2] Ren Y, Li Q, Du X, Zhang Y, Wang H, Shi G, Wei M. Analysis of corn yield prediction potential at various growth phases using a process-based model and deep learning. *Plants*. 2023;12(3):446. <https://doi.org/10.3390/plants12030446>
- [3] Bantchina BB, Qaswar M, Arslan S, Ulusoy Y, Gündoğdu KS, Tekin Y, Mouazen AM. Corn yield prediction in site-specific management zones using proximal soil sensing, remote sensing, and machine learning approach. *Computers and Electronics in Agriculture*. 2024;225:109329. <https://doi.org/10.1016/j.compag.2024.109329>
- [4] Kim N, Na S-I, Park C-W, Huh M, Oh J, Ha K-J, et al. An artificial intelligence approach to prediction of corn yields under extreme weather conditions using satellite and meteorological data. *Applied Sciences*. 2020;10(11):3785. <https://doi.org/10.3390/app10113785>
- [5] Kim N, Ha K-J, Park N-W, Cho J, Hong S, Lee Y-W. A comparison between major artificial intelligence models for crop yield prediction: Case study of the midwestern United States, 2006–2015. *ISPRS International Journal of Geo-Information*. 2019;8(5):240. <https://doi.org/10.3390/ijgi8050240>
- [6] HeidariGholanlo M, Javanmard Alitappeh R, Ghnabari Parmehr E. Large Scale Crop Type Mapping Using Multi Temporal Sentinel-2 Imagery and a New Attention U-Net-Vit. *Journal of Remote Sensing and Geoinformation Research*. 2024;2(2):233-46. [In Persian] <https://doi.org/10.22061/jrsgr.2024.11047.1071>
- [7] Terliksiz AS, Altılar DT. Impact of large kernel size on yield prediction: a case study of corn yield prediction with SEDLA in the US Corn Belt. *Environmental Research Communications*. 2024;6(2):025011. <https://doi.org/10.1088/2515-7620/ad27fa>
- [8] Ji Z, Pan Y, Zhu X, Zhang D, Dai J. Prediction of corn yield in the USA corn belt using satellite data and machine learning: from an evapotranspiration perspective. *Agriculture*. 2022;12(8):1263. <https://doi.org/10.3390/agriculture12081263>
- [9] Ju S, Lim H, Ma JW, Kim S, Lee K, Zhao S, Heo J. Optimal county-level crop yield prediction using MODIS-based variables and weather data: A comparative study on machine learning models. *Agricultural and Forest Meteorology*. 2021; 307: 108530. <https://doi.org/10.1016/j.agrformet.2021.108530>
- [10] Xiong D. Crop growth remote sensing monitoring and its application. *Sensors & Transducers*. 2014;169(4):174
- [11] Ma Y, Zhang Z, Kang Y, Özdoğan M. Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Remote Sensing of Environment*. 2021; 259: 112408. <https://doi.org/10.1016/j.rse.2021.112408>

معرفی نویسندگان

AUTHOR(S) BIOSKETCHES



علی سبزعلی یمقانی دکتری مهندسی نقشه برداری گرایش سیستم های اطلاعات مکانی از دانشگاه صنعتی خواجه نصیرالدین طوسی می باشد. ایشان مدرک کارشناسی ارشد و دکتری خود را در رشته مهندسی نقشه برداری- سیستم های اطلاعات مکانی در

سال های ۱۳۹۴ و ۱۳۹۸ از دانشگاه خواجه نصیرالدین طوسی دریافت نمودند. فعالیت های پژوهشی اخیر وی در حوزه شبکه های اجتماعی مکان مبنا، کاربرد یادگیری ماشین در پردازش تصاویر ماهواره ای و کاربردهای تصمیم گیری چندمعیاره می باشد. در حال حاضر ایشان بیش از ۱۵ مقاله علمی در مجلات و کنفرانس های علمی ارائه نموده اند و همچنین در کمیته علمی و داوری بیش از ۵ مجله و کنفرانس علمی ملی و بین المللی فعالیت داشته اند.

Sabzali Yameqani, A. Ph.D. at the Department of Geospatial Information Systems, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran

✉ asabzali@mail.kntu.ac.ir



علی اصغر آل شیخ استاد تمام گروه سیستم اطلاعات مکانی دانشگاه صنعتی خواجه نصیرالدین طوسی می باشد. ایشان دارای مدرک دکترای تخصصی در رشته مهندسی سیستم اطلاعات مکانی (GIS) از دانشگاه کلگری کانادا (۱۳۷۷) می باشند. همچنین،

جوایز و افتخارات متعددی برای مشارکت های صنعتی، علمی و دانشگاهی کسب کرده اند. مقالات بسیاری در مجلات معتبر خارجی و داخلی توسط ایشان در زمینه های مختلف همچون سیستم های اطلاعات مکانی، GIS تحت وب، مکان یابی، هوش مصنوعی، Health GIS و... به چاپ رسیده است.

Alesheikh, A. A. Full Professor at the Department of Geospatial Information Systems, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran

✉ alesheikh@kntu.ac.ir

for smart crop production prediction in the peasant farming domain. *Applied Sciences*. 2022;12(4):1940. <https://doi.org/10.3390/app12041940>

[24] Schmidt L, Odening M, Schlanstein J, Ritter M. Exploring the weather-yield nexus with artificial neural networks. *Agricultural Systems*. 2022;196:103345. <https://doi.org/10.1016/j.agsy.2021.103345>

[25] Van Klompenburg T, Kassahun A, Catal C. Crop yield prediction using machine learning: A systematic literature review. *Computers and electronics in agriculture*. 2020;177:105709. <https://doi.org/10.1016/j.compag.2020.105709>

[26] Meghraoui K, Sebari I, Pilz J, Ait El Kadi K, Bensiali S. Applied Deep Learning-Based Crop Yield Prediction: A Systematic Analysis of Current Developments and Potential Challenges. *Technologies*. 2024;12(4):43. <https://doi.org/10.3390/technologies12040043>

[27] Oikonomidis A, Catal C, Kassahun A. Hybrid deep learning-based models for crop yield prediction. *Applied artificial intelligence*. 2022;36(1):2031822. <https://doi.org/10.1080/08839514.2022.2031823>

[28] Luo S, Zhang S, Cong H, editors. Research on consumer purchasing prediction based on xgboost algorithm. 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA); 2021: IEEE.

[29] Trojovský P, Dehghani M. Pelican optimization algorithm: A novel nature-inspired algorithm for engineering applications. *Sensors*. 2022;22(3):855. <https://doi.org/10.3390/s22030855>

[30] Felegari S, Moravej K, Sharifi A, Golchin A, Karami P. Prediction of total soil nitrogen variations using three machine learning approaches and remote sensing data. 2023. [In Persian] <https://doi.org/10.22061/jrsgr.2024.10781.1060>

[31] Sabzali Yameqani A, Alesheikh AA. Predicting subjective measures of walkability index from objective measures using artificial neural networks. *Sustainable cities and society*. 2019;48:101560. <https://doi.org/10.1016/j.scs.2019.101560>

[32] Chen W, Zhang H, Mehlatat MK, Jia L. Mean-variance portfolio optimization using machine learning-based stock price prediction. *Applied Soft Computing*. 2021;100:106943. <https://doi.org/10.1016/j.asoc.2020.106943>

Citation (Vancouver): Sabzali Yameqani A, Alesheikh A. A. [Predicting Corn Yield Using Ensemble Learning with Satellite and Meteorological Data]. *J. RS. GEOINF. RES.* 2025; 3(1): 53-64

<https://doi.org/10.22061/jrsgr.2025.11691.1092>



COPYRIGHTS

© 2025 The Author(s). This is an open-access article distributed under the terms and conditions of the Creative Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) (<https://creativecommons.org/licenses/by-nc/4.0/>)